Mon. Not. R. Astron. Soc. **000**, 1–10 (2011) Printed 21 January 2011 (MN LATEX style file v2.2)

Connected structure in the 2dFGRS

D.N.A. Murphy⋆, V.R. Eke and Carlos S. Frenk
Institute for Computational Cosmology, Department of Physics, University of Durham, South Road, Durham, DH1 3LE, UK

ABSTRACT
We describe and apply a simple prescription for defining connected structures in galaxy redshift surveys. The method is based upon two passes with a friends-of-friends groupfinder. The first pass uses a cylindrical linking volume to find galaxy groups and clusters, in order to suppress the line-of-sight smearing introduced by the large random velocities of galaxies within these deep potential wells. The second pass, performed with a spherical linking volume, identifies the connected components. This algorithm has been applied to the 2dFGRS, within which it picks out a total of 7,603 systems containing at least two galaxies and having a mean redshift less than 0.12. Connected systems with many members appear filamentary in nature, and the algorithm recovers two particularly large filaments within the 2dFGRS. For comparison, the algorithm is has also been applied to ΛCDM mock galaxy surveys. While the model population of such systems is broadly similar to that in the 2dFGRS, it does not generally contain such extremely large structures.

Key words: cosmology: observations – large-scale structure of universe

1 INTRODUCTION

At very large scales, baryonic material is concentrated into an interconnected sponge-like network known as the Cosmic Web (Bond, Kofman & Pogosyan 1996). Successive redshift surveys have traced out imposing overdensities in the galaxy distribution. Notable examples are the CfA 'Great Wall' (Geller & Huchra 1989) and the 'Sloan Great Wall' (Gott et al. 2005), which was also noted, if not named, in the Two-degree Field Galaxy Redshift Survey (2dFGRS, Colless et al. 2001) by Baugh et al. (2004) and Erdoğdu et al. (2004).

Most studies of large-scale structure in the Universe concentrate on measuring the galaxy power spectrum (e.g. Cole et al. 2005) or its Fourier transform, the two-point correlation function (e.g. Zehavi et al. 2002). These quantities will provide a complete statistical description of the galaxy distribution provided that their number density fluctuations are Gaussian. The standard model of structure formation, ΛCDM, does assume that the initial inhomogeneities in the density field, generated during inflation, are Gaussian. However, the subsequent growth of structure due to gravitational instability induces significant non-Gaussianities in the density field, and higher order moments of the density distribution become important. These phase correlations within the density field can be characterised either through higher order galaxy correlation functions (Baugh et al. 2004; Croton et al. 2004; Gaztañaga et al. 2005; Nichol et al. 2006) or through the properties and distribution of filaments. Quantitative studies of filamentary structures in redshift surveys have been performed by Bhavsar & Barrow (1983); Barrow, Bhavsar & Sonoda (1985); Einasto et al. (2003); Pimbblet & Drinkwater (2004) and Stoica, Martínez & Saar (2010).

A variety of algorithms have been designed to describe the morphology of these large-scale structures using different techniques such as percolation (Bhavsar & Barrow 1983), visual identification of regions between clusters (Pimbblet, Drinkwater & Hawkrigg 2004; Colberg, Krughoff & Connolly 2005), minimal spanning trees (Barrow, Bhavsar & Sonoda 1985; Colberg 2007), the density field's Hessian (Aragón-Calvo et al. 2007; Bond, Strauss & Cen 2010; Zhang et al. 2009), gradient (Morse theory, Novikov, Colombi & Doré 2006) or linkage between the two (Sousbie et al. 2008a,b), the Hessian of the potential field (Hahn et al. 2007; Forero-Romero et al. 2009), Delaunay tessellations (Schaap & van de Weygaert 2000; Aragón-Calvo et al. 2007), the Candy and Bisous models (Stoica et al. 2005; Stoica, Martínez & Saar 2010) and watershed transforms (Sousbie, Colombi & Pichon 2009; Aragón-Calvo et al. 2010). Many of these algorithms also partition the whole of space into clusters, walls, filaments and voids. They are often applied to dark matter simulations to help describe the mass distribution, but with a few notable exceptions (Bhavsar & Barrow 1983; Barrow, Bhavsar & Sonoda 1985; Pimbblet & Drinkwater 2004; Bond, Strauss & Cen 2010; Stoica, Martínez & Saar 2010), they rarely include a comparison with observational data. A primary motivation for this paper is to carry out a detailed quantitative comparison of filament properties using the 2dFGRS and mock galaxy catalogues created us-

⋆ david.murphy@durham.ac.uk

ing a ΛCDM simulation (Angulo et al. 2008) and a semi-analytical galaxy formation model (Baugh et al. 2005).

An important aspect of the comparison between model and observed large-scale structure concerns the existence in both the 2dFGRS and the Sloan Digital Sky Survey (SDSS; York et al. 2000) of some extremely large structures. These objects are known to have a large impact on the higher order correlations of the galaxy distribution (Baugh et al. 2004; Croton et al. 2004; Gaztañaga et al. 2005; Nichol et al. 2006), and on its topology (Park et al. 2005). How common such structures are within the ΛCDM model remains contentious (Yaryura, Baugh & Angulo 2010).

In Section 2, we describe the detection algorithm that we have developed, and the observational and mock data that will be compared. The results of this comparison are described in Section 3 and conclusions drawn in Section 4.

2 METHODS

In this section, we will describe the observational data that will be analysed, the algorithm for finding connected systems, the calculation of the luminosity of such objects, and the mock catalogues that will be used in order to compare the ΛCDM model with the observational data.

2.1 Data

The observational data used in this study are the two large contiguous patches towards the north and south galactic poles (NGP and SGP) in the final data release of the 2dFGRS (Colless et al. 2001). In total, these regions contain 191,897 galaxies with a median redshift of 0.11, covering an area of approximately 1500 deg^2 to a flux limit corresponding to $b_J \sim 19.45$. A bright flux limit of $b_J < 14$ is also imposed to exclude objects whose total fluxes are difficult to determine from APM photographic plates.

2.2 The detection algorithm

A desirable property of an algorithm extracting large-scale structure is that there should be no preferred direction for the resultant systems. However, redshift space distortions make the line of sight a special direction in galaxy redshift surveys. The most striking consequence of non-Hubble flow velocities is to stretch galaxy clusters, creating 'fingers of god' (Jackson 1972) in the redshift-space galaxy distribution. These elongated redshift-space distortions need to be removed before searching for real structures. We achieve this by taking the 2PIGG catalogue of groups and clusters constructed by Eke et al. (2004a) from the 2dFGRS. These were found using a friends-of-friends algorithm with a cylindrical linking volume pointing along the redshift direction as described in their paper. Having found galaxies belonging to groups and clusters in this way, we would like to collapse the 'fingers of god' by placing these galaxies at their group centre positions. One complication is that Eke et al. (2004a) note that they would expect the 2PIGG catalogue to contain a few tens of per cent of interloper galaxies that are incorrectly assigned to groups. To try and correct for this inevitable misassignment, we choose to retain a redshift-dependent fraction

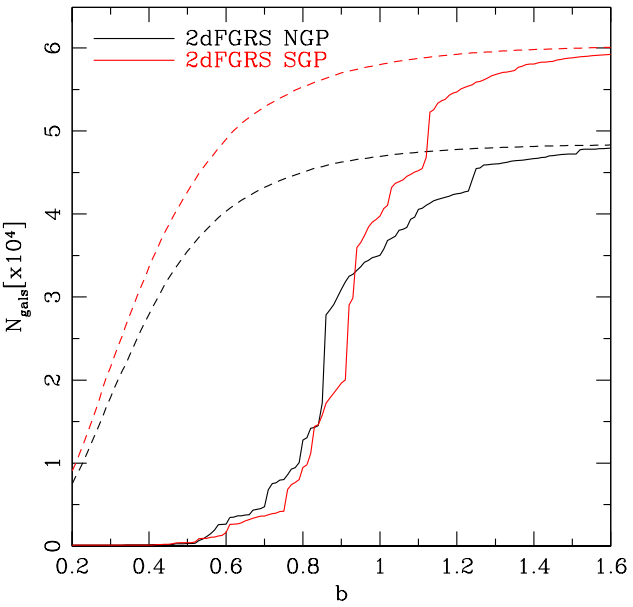

Figure 1. Variation in the number of 2dFGRS galaxies in systems extracted by the algorithm as the relative linking length, b, changes. Black lines denote galaxies from the NGP, red lines from the SGP. Solid lines represent the number of galaxies in the largest structure, whilst the dashed lines show the number of galaxies in all systems with at least two members.

$$f(z) = \frac{2-z}{2.4+z} \quad (1)$$

of the members assigned to each group. This expression for $f(z)$ is derived from the contamination of groups found in mock 2dFGRS catalogues by Eke et al. (2004a). The randomly selected fraction $f(z)$ are all replaced by a single point at the group centre, whereas the $1-f(z)$ of 'interloper' galaxies are jettisoned from the list of group members and left at their observed redshift space positions.

The first friends-of-friends pass suppresses the redshift space distortions associated with intragroup line-of-sight galaxy velocities. Note that this collapse does not account for the coherent infall of galaxies onto overdensities that will enhance and merge structure in the plane of the sky (see e.g, Kaiser 1987; Praton, Melott & McKee 1997). We then apply a friends-of-friends algorithm with a spherical linking volume to the set of remaining galaxies and group centres. The radius of this linking sphere is chosen to be b times the mean intergalactic separation at that redshift, as defined in Eqn. 2.7 of Eke et al. (2004a). Small linking lengths would lead to many small systems, whereas very large ones would lead to percolation, and a single large connected component encompassing everything in the survey. An intermediate value for b will lead to a more useful description of the structures present in the survey. This two-pass procedure provides a new and simple way to define connected structures in galaxy redshift surveys.

Fig. 1 shows how the number of galaxies in connected structures and the number of galaxies in the largest system vary with b for both the NGP and SGP. Both the NGP and SGP regions show a rapid growth of their largest system as b increases beyond ~ 0.8. Fig. 2 shows the variation of the

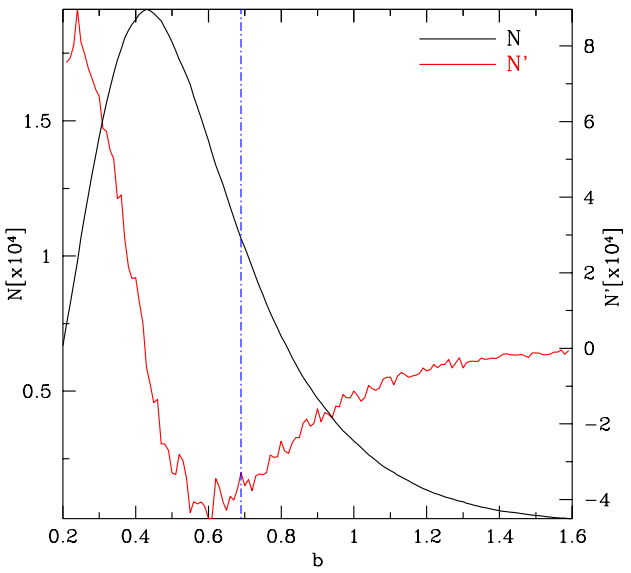

Figure 2. Variation in the number of structures extracted by the algorithm as the relative linking length, b, varies. The red line represents the first derivative of this function, corresponding to the rate of change of system number. We adopt a relative linking length, $b = 0.69$, close to this minimum, as denoted by the dot-dashed blue line. This corresponds to systems approximately bounded by a surface with a galaxy number overdensity of ~ 1.5.

total number of connected structures and its first derivative. We pick $b = 0.69$ as a value that gives rise to an interestingly large range of system sizes. This corresponds to finding structures bounded by an irregular surface that has an overdensity of $\Delta\rho/\bar{\rho} = 3/(2\pi b^3) \approx 1.5$ (Cole & Lacey 1996).

This choice is close to the point at which $\dot{N}/\dot{b} = 0$, where the growth in the number of systems arising from single galaxies becoming linked matches the decrease caused by merging the structures together. The resulting systems are shown in Fig. 3.

The abundance and extent of survey-sized connected structures will depend upon the geometry of the survey to which this algorithm is applied. Thus, while this technique is appropriate for comparing an observed data set with a mock catalogue of that particular survey, care is required when trying to infer the physical properties and abundance of the largest structures in the underlying distribution.

2.3 Connected structure luminosities

We would like to quantify the sizes of the objects found using this method in a way that (a) does not depend explicitly on the magnitude limit of the survey and (b) assigns the same size to a particular structure, independently of the redshift at which it is placed. Thus, rather than merely counting the number of galaxies present in each system, we define a luminosity that takes into account the flux limits of the survey. The angular variation of the flux limit in the 2dFGRS is such that it changes over the length of the elongated filamentary structures. Consequently, it is necessary to convert the observed luminosity of each galaxy to the total luminosity that would have been seen without any flux limits, rather than correcting the observed luminosity of the system as a whole. This is done assuming that the galaxy luminosity function $\Phi(L)$ is given by the Schechter function determined by Norberg et al. (2002), i.e. $(L_*, \alpha) = (10^{10}\, h^{-2}\, \mathrm{L}_\odot, -1.21)$ and using

$$\frac{L_{\rm cor}}{L_{\rm obs}} = \frac{\int_0^\infty L\Phi(L)dL}{\int_{L_{min}}^{L_{max}} L\Phi(L)dL}, \qquad (2)$$

where the luminosity limits in the integral in the denominator reflect the upper and lower flux limits evaluated at the redshift of the galaxy. $L_{\rm cor}$ and $L_{\rm obs}$ represent the corrected and observed galaxy luminosities respectively, taking into account the $k + e$ correction in a manner similar to Norberg et al. (2002):

$$k + e = \frac{z + 6z^2}{1 + 8.9z^{2.5}}. \qquad (3)$$

The total luminosity is calculated by summing up all the corrected galaxy luminosities for galaxies within that system, taking into account the weighting factors that describe the local incompleteness of the survey. Given the flux limit of the 2dFGRS, the fraction of the total luminosity that is observed drops beneath a half at redshifts exceeding $z \sim 0.12$. For this reason, we will restrict our analysis to structures with $\bar{z} \leqslant 0.12$. Fig. 4 shows the systems found within the 2dFGRS colour coded according to their luminosity. Comparing with Fig. 3, it is apparent how the luminosity picks out structures at larger distances than the membership, which includes only galaxies within the flux limits.

2.4 Mock catalogues

In order to address how well the observed distribution of system luminosities compares with that predicted for a 2dFGRS-like survey of a ΛCDM cosmological model, we need to create mock galaxy surveys. This is done using a combination of the BASICC dark matter simulation described by Angulo et al. (2008), and the semi-analytical model of Baugh et al. (2005), which is a development of that introduced by Cole et al. (2000).

In brief, the BASICC simulation contains 1448^3 particles in a $1340h^{-1}$ Mpc-long cube of a ΛCDM model with $\Omega_{\rm M} = 0.25$, $\Omega_\Lambda = 0.75$, $\Omega_{\rm b} = 0.045$, $h = 0.73$ and $\sigma_8 = 0.9$. Angulo et al. (2008) used a friends-of-friends (FOF) algorithm with a linking length of 0.2 times the mean interparticle separation to define haloes. Requiring 10 particles to resolve a halo implies that the minimum resolvable halo mass is $5.5 \times 10^{11} h^{-1}\, \mathrm{M}_\odot$. All of these haloes were populated with galaxies according to the semi-analytical model of Baugh et al. (2005). Galaxies that reside in unresolved haloes are randomly placed onto dark matter particles outside resolved haloes, according to the method described by Cole et al. (2005).

While the galaxy luminosity function produced by this model is close (within 0.3 magnitudes around L_*) to that observed in the 2dFGRS Norberg et al. (2002), we nevertheless apply a luminosity-dependent shift in luminosities so that the cube of model galaxies has exactly the same luminosity function as the 2dFGRS. Only galaxies with

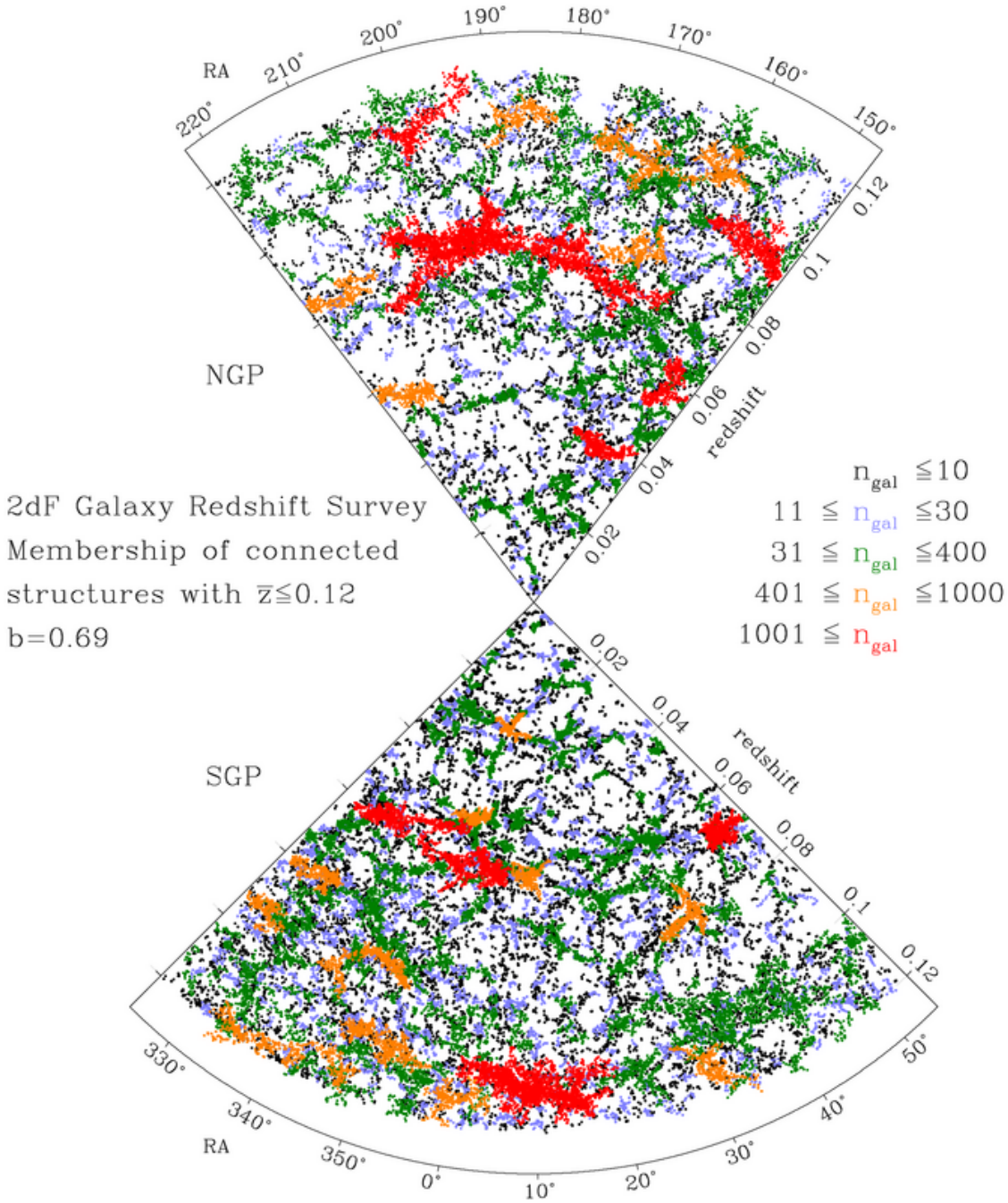

Figure 3. Spatial distribution of 2dFGRS galaxies in connected structures for systems with average redshift $\bar{z} \leqslant 0.12$ in the RA-z plane. These systems contain at least two galaxies and dot colours represent the weighted number of galaxies in the structure, where this weight takes into account the local angular incompleteness.

$L > 1.4 \times 10^8\, h^{-2}\, \mathrm{L}_\odot$ are included in the model cube. For the flux limit of the 2dFGRS, this implies that the mock galaxy catalogues will be missing low luminosity galaxies at $z \lesssim 0.025$. This corresponds to just under one per cent of the total volume being considered.

Fifty mocks of the 2dFGRS were created from this cube of model galaxies as follows:
a) a random observer location and direction were chosen,
b) volume-limited NGP and SGP surveys were created using periodic replicas of the cube if required. Note that the depth of the 2dFGRS is less than half the length of the BASICC simulation cube, and its effective volume is less than 10^{-3} of the volume of the BASICC simulation,
c) galaxies were removed according to the position-dependent 2dFGRS flux limits and completeness masks,
d) the remaining galaxies were assigned a redshift according to their peculiar velocity in the simulation.

Previous studies have shown that this semi-analytic model tends to place too many low luminosity galaxies into galaxy clusters (Eke et al. 2004b; Gilbank & Balogh 2008; Kim et al. 2009). Given that the global luminosity function

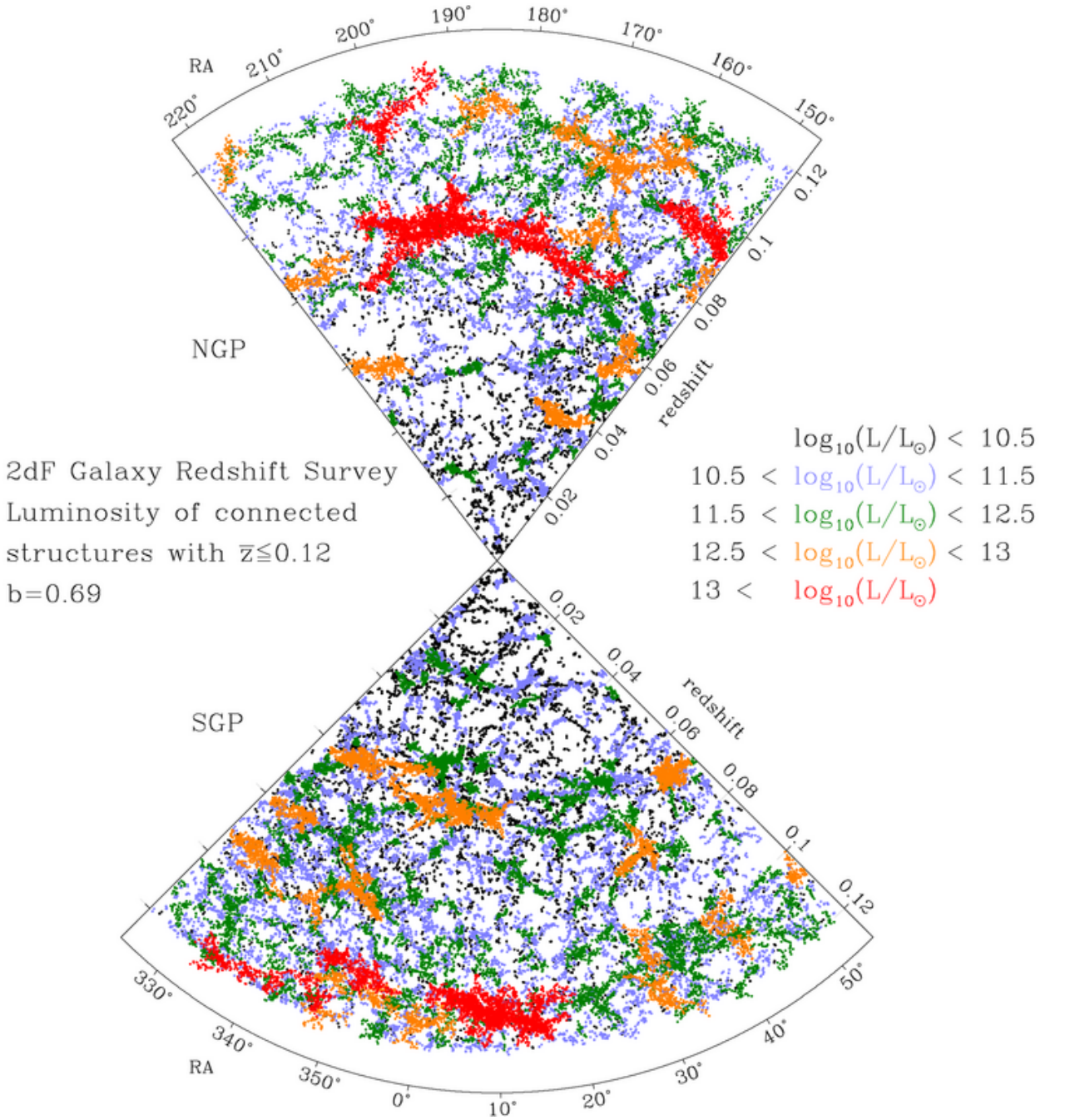

Figure 4. Spatial distribution of 2dFGRS galaxies in connected structures in the RA-z plane. Colours represent the total system luminosity in units of $\log_{10}(L/h^{-2}L_{\odot})$.

has been forced to match that in the observations, this implies that the model will lack low luminosity galaxies in lower density regions. This known problem will affect the structure finder.

To try and reduce the impact of this known difference between the model and observations, we allow ourselves the freedom to jettison a smaller fraction of galaxies from the model groups than given by Eqn. 1 for the real 2dFGRS. This decreases, in the vicinity of the groups, the number density of points used for the structure-finding sweep of the friends-of-friends algorithm to an amount similar to that in the real survey. We achieve this in the model by multiplying $f(z)$, as given by Eqn. 1, by a constant, χ. $\chi > 1$ implies that a higher fraction of galaxies are retained in the groups.

In order to determine an appropriate value for χ, we have measured the distribution of system orientations defined as

$$\theta = \tan^{-1}\left(\frac{\Delta l_z}{\Delta l_\phi}\right), \qquad (4)$$

where Δl_z represents the range of the member galaxies in the redshift direction and Δl_ϕ is the larger of the ranges of member galaxies in the RA and dec directions. Thus,

$\theta = \pi/2$ for a radial object and 0 for one lying perpendicular to this. We use the greater of Δl_z and Δl_ϕ to describe the scale size of the connected structure.

Fig. 5 shows the cumulative probability distributions of system orientations for structures containing at least 20 galaxies in the 2dFGRS and those recovered from 10 mock surveys using three different values of χ[1]. It is apparent that, when treated in the same way as the real data (i.e. $\chi = 1$), the model contains too many objects aligned along the line of sight. This is a result of too many low luminosity galaxies being placed into the redshift space volumes occupied by the model groups. When the 'interloping' galaxies are jettisoned from the groups found in the mock catalogues, enough of these extra low luminosity galaxies are placed along radial lines that they bias the orientation distribution. Increasing χ retains a higher fraction of the initially grouped galaxies in the groups, reducing the number of 'interlopers' returned to the field, and decreasing the number of radially aligned objects found in the second pass of the friends-of-friends algorithm. A value of $\chi = 1.15$ produces a mock orientation distribution that is, according to a K-S test, indistinguishable from that found in the 2dFGRS. This is chosen as the default value for these 2BASICC mocks throughout this paper.

An additional set of 22 mock 2dFGRS surveys created from the Hubble Volume by Cole et al. (2000), as used by Gaztañaga et al. (2005), are also analysed to give some idea of the systematic differences resulting from different simulations and implementations of the galaxy formation modelling. For these mock catalogues, a value of $\chi = 1.11$ was required in order to recover the 2dFGRS system orientation distribution.

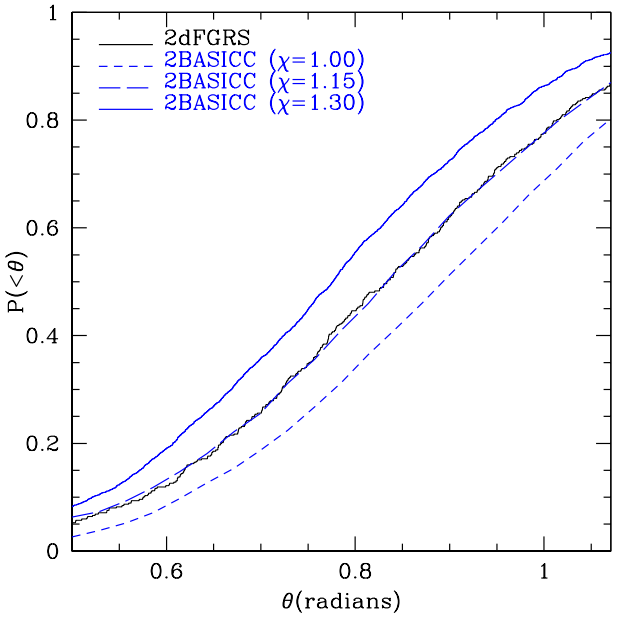

Figure 5. Cumulative probability distributions of system orientations for all objects containing at least 20 galaxies. Results are shown for the 2dFGRS and for averages of 10 2BASICC mocks. The mock distributions are derived from three different choices of $\chi = 1.0, 1.15$ and 1.30, as indicated in the legend.

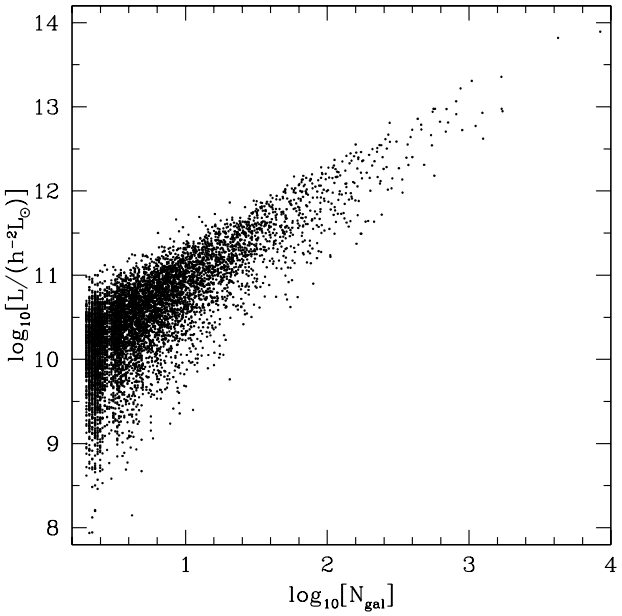

Figure 6. The relation between object luminosity, L, and N_{gal}, the weighted number of galaxies in objects with $\bar{z} \leqslant 0.12$.

3 RESULTS

In this section, we will first describe the main features and properties of the connected structures found in the 2dFGRS and then compare with the results from the ΛCDM mock surveys. This comparison will encompass both the whole population of connected systems as well as the largest objects.

3.1 Connected systems in the 2dFGRS

Projections onto the right ascension-redshift plane of all of the connected systems found in the 2dFGRS are shown in Figs.3 and 4. A total of 95,010 galaxies are linked into 7,603 systems containing at least two members and mean redshifts no greater than 0.12. Of these, 3,018 contain only two members. Almost 87 per cent of galaxies at $z \leqslant 0.12$ are placed into a connected structure. One large filamentary-structure stands out in each of the NGP and SGP wedges. These systems trace out the same overdensities apparent in the 2PIGG distribution (Eke et al. 2004b), the smoothed galaxy density map (Baugh et al. 2004) and the reconstructed density field (Erdoğdu et al. 2004) of the 2dFGRS. The largest NGP object, at $z \sim 0.08$, corresponds to the large RA end of the 'Sloan Great Wall' highlighted by Gott et al. (2005). At a total b_J-band luminosity of $\sim 7.8 \times 10^{13}\, h^{-2}\, L_\odot$, this is about 20 per cent more luminous than the largest equivalent in the SGP, which lies at $z \sim 0.11$ and RA $\sim 10^\circ$.

[1] One might imagine that randomly oriented connected structures would be uniformly distributed with θ. However, since systems often contain more than two galaxies, which are generally not colinear, the definition of θ leads to connected structures preferentially avoiding values towards the ends of the range $[0, \pi/2]$.

Table 1. Properties of connected structures with $\bar{z} \leqslant 0.12$ identified in the 2dFGRS and mock surveys. Mock values are the mean over all 50 2BASICC and 22 Hubble Volume mock surveys with the uncertainties being the standard deviation of individual surveys from these mean values. N is the total number of connected systems within the catalogue, f the fraction of galaxies out to $z = 0.12$ in systems and $N_{\rm gal}$ is the total number of galaxies out to $z = 0.12$. The fifth and sixth columns describe the average and maximum object luminosities (in units $\log_{10}(L/h^{-2}L_{\odot})$). We give the comoving scale l_{max} of the largest structure identified in the survey in the final column. This is defined as the largest in extent of galaxy members in the redshift, RA or dec directions.

Survey	N	f	$\log_{10}(N_{\rm gal})$	$\log_{10}\bar{L}$	$\log_{10} L_{max}$	$l_{max}(h^{-1}\,{\rm Mpc})$
2dFGRS	7603	87.7%	5.06	11.16	13.89	198
2BASICC	8023 ± 250	$85.9 \pm 0.8\%$	5.07 ± 0.04	11.08 ± 0.03	$13.44^{+0.15}_{-0.23}$	81 ± 19
HV	8253 ± 135	$82.0 \pm 0.8\%$	5.11 ± 0.02	11.10 ± 0.03	$13.55^{+0.14}_{-0.20}$	93 ± 27

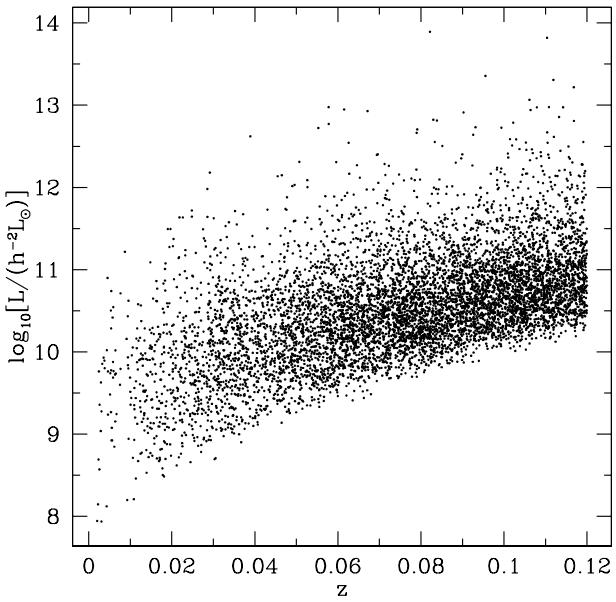

Figure 7. The distribution of system luminosities with increasing redshift.

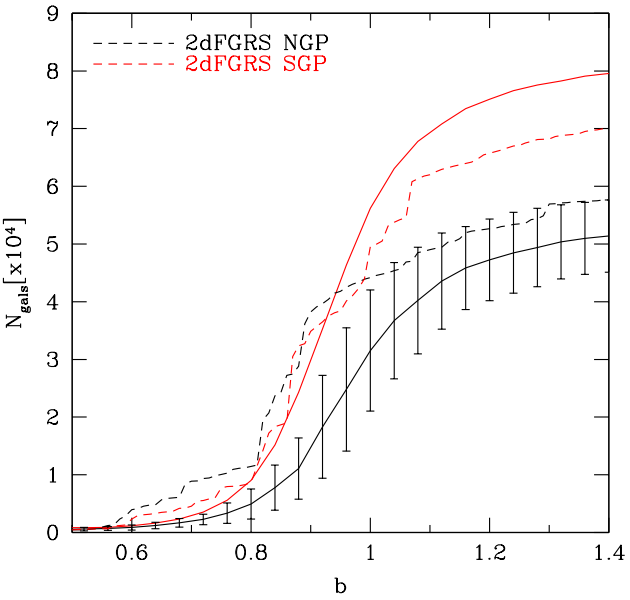

Figure 8. The weighted number of galaxies in the largest structure for NGP (black) and SGP (red) systems with $\bar{z} \leqslant 0.12$, including any members with redshifts greater than this limit subject to the mean system redshift remaining below it. In both cases the dashed lines show the 2dFGRS data, whilst solid lines represent the mean number of galaxies in the largest object across 50 mock surveys. In the NGP case, we include also error bars representing the standard deviation of these surveys around the mean.

The extents in RA of these largest NGP and SGP systems in comoving coordinates are $\sim 198h^{-1}$ Mpc and $99h^{-1}$ Mpc respectively. While the NGP systems contains twice as many members as that in the SGP, it is very nearly broken into two pieces around RA $\sim 185^{\circ}$, where the galaxy density drops off considerably.

More locally, a continuation to lower declinations of the CfA Great Wall (Geller & Huchra 1989) is seen at $z \sim 0.02$ in the NGP, although our algorithm breaks this up into a few different components.

Some average and extreme properties of the systems identified in the 2dFGRS are listed in Table 1. In more detail, the correlation between the luminosity and weighted (to account for the local angular incompleteness of the survey) membership of connected structures is shown in Fig. 6. The second largest system contains at least twice as many members as the third largest one, and almost 3 times as much luminosity, making the largest NGP and SGP structures stand out from the remaining systems. The scatter around the mean relation reflects the range of redshifts in the flux-limited survey. While the object luminosity is corrected to take account of this flux limit, the weighted number of galaxies is not.

Fig. 7 shows how connected system luminosity varies with redshift, with the lower envelope representing the total luminosity of 2 galaxies at the flux limit. The geometry of the survey precludes finding very luminous structures at low redshift because of the small volume sampled, but the greater volume available at larger redshifts is sufficient to contain larger, more luminous filamentary structures. The largest NGP and SGP systems are once again conspicuous at the top of the figure.

3.2 Comparison with mock catalogues

We follow almost the same procedure to define connected structures in the 50 2BASICC mocks, with the only difference being that the fraction of galaxies retained in the groups, $f(z)$ in Eqn. 1, is increased by a factor $\chi = 1.15$, as

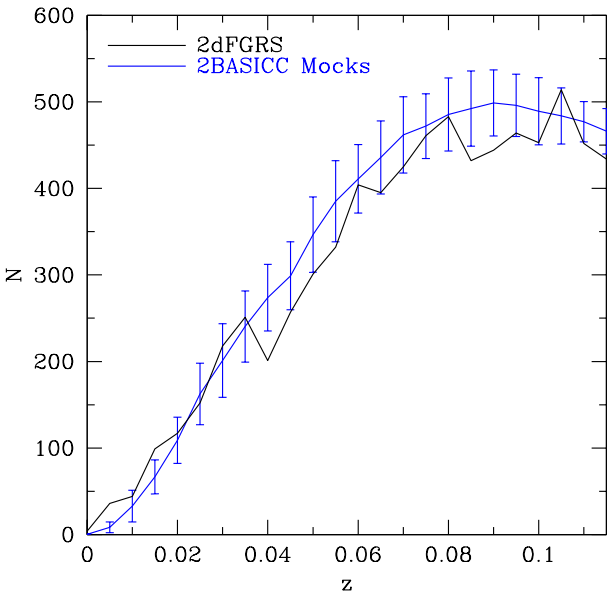

Figure 9. The redshift distribution of all objects with at least two members. The blue line represents the mean number of connected structures as a function of redshift across 50 mock surveys, with error bars representing their standard deviation around the mean. The black line corresponds to systems detected in the 2dFGRS.

described in §2. The impact of this choice on the results is discussed later.

Fig. 8 shows how the number of galaxies in the most populated system grows as b is increased for each survey region. While the behaviour is broadly similar to that of the largest filament in the 2dFGRS, the onset of percolation in mock catalogues is delayed by about 0.1 times the mean intergalaxy separation. As a consequence, at $b = 0.69$, the largest 2dFGRS system (located in the NGP) is a significant outlier, being more populated than the corresponding structure in all but one of the 50 mock catalogues. The mock system containing more galaxies than the largest one in the 2dFGRS is placed at $z \sim 0.01$ and is much less luminous. It is the result of the randomly chosen observer being put very near to a large galaxy cluster.

The larger value of χ adopted for mock surveys means that the number of galaxies and group centres used for the algorithm is on average $\sim$ 13 per cent lower than in the case of the 2dFGRS. If we were to use a value of b that was correspondingly larger (i.e. scaled by the inverse cube root of the number of points to be 0.72) then this does not significantly affect the discrepancy between the number of galaxies in the most populated systems in the mock or real 2dFGRS.

The redshift distribution of the structures is shown in Fig. 9. At $z \lesssim 0.025$, where the mock catalogues are known to be missing low luminosity galaxies, the mock surveys contain fewer objects than the real 2dFGRS. The main reason for this is actually not the incompleteness in the mocks, but the fact that too many low luminosity galaxies are placed into large groups, reducing the number available to form other small systems. This local volume represents only a small fraction of the survey.

For redshifts greater than 0.04, the number of real 2dFGRS structures is typically slightly below the mean of the 50 mock surveys. This is reflected in the first column of Table 1, which shows that the total number of systems in the 2dFGRS is $(1-2)\sigma$ beneath that of the mocks, despite the fact that a slightly higher fraction of galaxies are placed into the 2dFGRS structures. The total number of galaxies in the 2dFGRS and mock surveys matches well by construction, but the excess of mock galaxies placed into groups means that fewer are available in lower density regions for linking together small systems. The relatively high fraction of galaxies placed into structures and low number of structures in the 2dFGRS leads to a larger mean luminosity. This difference can be removed by not including the two most luminous systems in the 2dFGRS in the calculation.

The distribution of system luminosities is shown in Fig. 10. The mocks have a relative lack of structures beneath $L \sim 10^9\, h^{-2}\, \mathrm{L}_\odot$, more than the real 2dFGRS at $L \sim 3 \times 10^{10}\, h^{-2}\, \mathrm{L}_\odot$, which is the peak of the distribution and corresponds approximately to two L_* galaxies, and a paucity of filamentary systems like the largest ones in the 2dFGRS. As stated above, the difference between the model and real distributions at low luminosities arises mostly because the lowest luminosity galaxies in the model are more likely to be placed into larger groups and hence are not available to form very low luminosity systems. The deficit of lower luminosity galaxies outside groups impacts in two ways upon the most luminous model structures. They tend to gain luminosity because their groups are slightly more luminous than those of corresponding mass in the 2dFGRS. However, the lack of low luminosity galaxies in the lower density regions, makes it less likely that large structures will join together. It is this second effect that is more important, resulting in none of the 50 mock surveys yielding two systems as luminous as the two most luminous in the 2dFGRS.

Given that many differences between the real and mock structure luminosity distributions result from the different spatial distributions of low luminosity galaxies in the real and mock surveys, and that we have used a different value of χ for real and mock surveys, one might reasonably ask what changes when $\chi = 1$ is used for the mocks. This is shown in Fig. 11, where three different χ values are used for the 2BASICC mocks. Increasing χ retains more galaxies in the groups, leaving fewer galaxies to help the algorithm link together larger structures. This leads to a decrease in the luminosity of the most luminous systems. Decreasing χ leads to an increase in the luminosity of the most luminous systems, but even for $\chi = 1$ there are still no surveys with two structures at least as luminous as the second most luminous 2dFGRS system. Nevertheless, we do obtain two surveys with a connected system more luminous than the brightest 2dFGRS system. However, as shown in Fig. 5, this comes at the expense of producing a set of objects that are significantly more radially oriented than those found in the 2dFGRS.

Also shown in Fig. 11 is the distribution of systems found in the 22 Hubble Volume mock catalogues of Cole et al. (2000). After tuning χ to be 1.11 to recover the 2dFGRS structure orientation distribution, the Hubble Volume mocks are broadly similar to the 2BASICC ones, with the abundance of the most luminous systems being almost

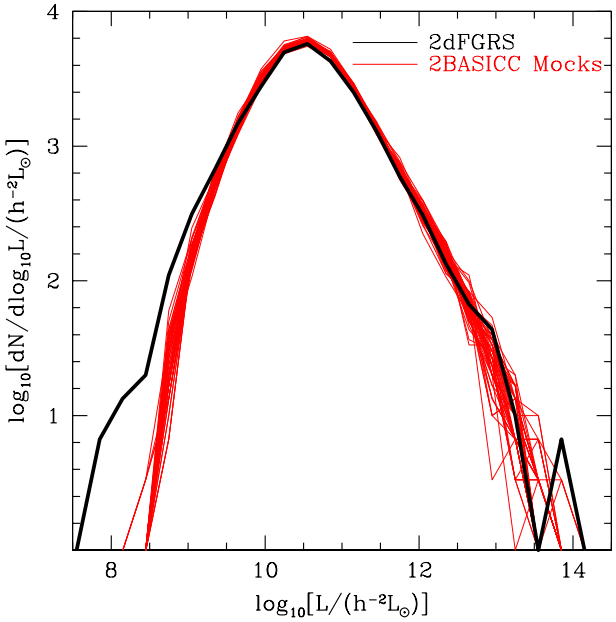

Figure 10. The distribution of luminosities for all structures with a minimum membership of two, out to a redshift $z = 0.12$. The red lines show the distribution for each of the 50 2BASICC mock catalogues. The black line shows the distribution for systems in the 2dFGRS.

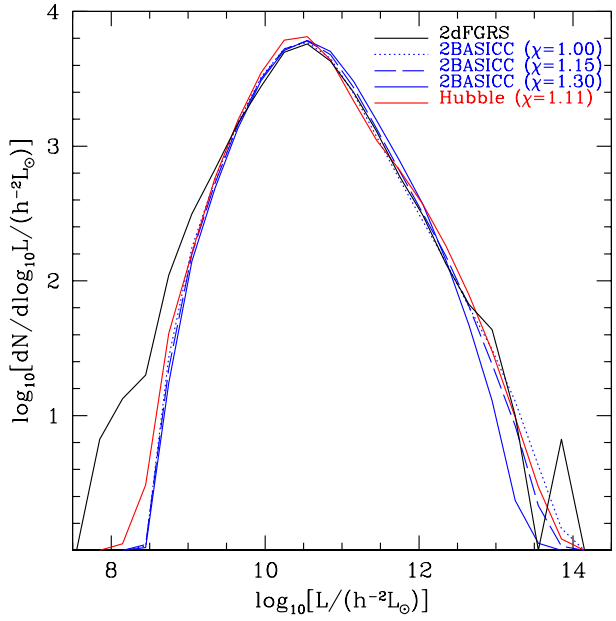

Figure 11. Structure luminosity distributions for different values of χ in the 2BASICC surveys, for the Hubble Volume catalogues and the 2dFGRS. Mock survey distributions have been averaged over the 50 surveys in the 2BASICC and the 22 in the Hubble Volume simulations.

unchanged. The bottom row of table 1 contains statistics for the systems found in these Hubble Volume mocks.

4 CONCLUSIONS

We have described a simple algorithm with which to define connected structure within galaxy redshift surveys, and applied it to the 2dFGRS. This algorithm explicitly addresses the redshift-space distortion associated with rapidly moving galaxies within groups and clusters. The 7,603 2dFGRS connected structures at $z \leqslant 0.12$ containing at least two members range up to $\sim 200h^{-1}$ Mpc in extent, but are mostly associations of two L_* galaxies. Quantifying object sizes via their total luminosities, corrected for the survey flux limits, we find that the largest systems are filamentary in nature and have b_J luminosities of almost $10^{14}\,h^{-2}\,\mathrm{L}_\odot$.

Applying the same algorithm to ΛCDM mock 2dFGRS catalogues, constructed using large-volume dark matter simulations and the semi-analytical model of Baugh et al. (2005), we find a broadly similar distribution of structures to those in the real data. There are, however, a few differences in detail. Many of these result from the fact that the model places too many $L \lesssim L_*$ galaxies into groups and clusters compared with the 2dFGRS. This biases the orientation distribution of the systems containing at least 20 galaxies to contain more radially aligned objects in the mock survey than in the 2dFGRS. Applying a crude correction to the algorithm to enable it to recover the same orientation distribution in the mock survey as it does in the 2dFGRS leads to the largest mock structures being significantly less luminous than those in the 2dFGRS.

It is clear that at least some of the differences between the properties of the structures in the 2dFGRS and the mock catalogues arise from inadequacies in the galaxy formation model that was used to construct the mocks. We have attempted to overcome these inadequacies as far as possible through empirical corrections. Our analysis indicates that the largest filamentary structures seen in the 2dFGRS are not reproduced in the mock catalogues. However, while this discrepancy could signal a failure of the standard ΛCDM cosmological model on large scales, it seems more plausible that it reflects a shortcoming in the predictions of our models of galaxy formation for the abundance and spatial distribution of galaxies on small scales.

ACKNOWLEDGEMENTS

DNAM acknowledges an STFC PhD studentship and CSF a Royal Society Wolfson Research Merit award. We thank Raul Angulo and Carlton Baugh for assistance in the production of the 2BASICC mock surveys. We also thank Kevin Pimblett and the referee for useful comments that improved the quality of this manuscript.

REFERENCES

Angulo R. E., Baugh C. M., Frenk C. S., Lacey C. G., 2008, MNRAS, 383, 755

Aragón-Calvo M. A., Jones B. J. T., van de Weygaert R., van der Hulst J. M., 2007, A&A, 474, 315

Aragón-Calvo M. A., Platen E., van de Weygaert R., Szalay A. S., 2010, ApJ, 723, 364
Barrow J. D., Bhavsar S. P., Sonoda D. H., 1985, MNRAS, 216, 17
Baugh C. M., Croton D. J., Gaztañaga E., Norberg P., Colless M., Baldry I. K., Bland-Hawthorn J., Bridges, T. et al., 2004, MNRAS, 351, L44
Baugh C. M., Lacey C. G., Frenk C. S., Granato G. L., Silva L., Bressan A., Benson A. J., Cole S., 2005, MNRAS, 356, 1191
Bhavsar S. P., Barrow J. D., 1983, MNRAS, 205, 61P
Bond J. R., Kofman L., Pogosyan D., 1996, Nature, 380, 603
Bond N. A., Strauss M. A., Cen R., 2010, MNRAS, 406, 1609
Colberg J. M., 2007, MNRAS, 375, 337
Colberg J. M., Krughoff K. S., Connolly A. J., 2005, MNRAS, 359, 272
Cole S., Lacey C., 1996, MNRAS, 281, 716
Cole S., Lacey C. G., Baugh C. M., Frenk C. S., 2000, MNRAS, 319, 168
Cole S., Percival W. J., Peacock J. A., Norberg P., Baugh C. M., Frenk C. S., Baldry I., Bland-Hawthorn, J. et al., 2005, MNRAS, 362, 505
Colless M., Dalton G., Maddox S., Sutherland W., Norberg P., Cole S., Bland-Hawthorn J., Bridges, T. et al., 2001, MNRAS, 328, 1039
Croton D. J., Gaztañaga E., Baugh C. M., Norberg P., Colless M., Baldry I. K., Bland-Hawthorn J., Bridges, T. et al., 2004, MNRAS, 352, 1232
Einasto J., Hütsi G., Einasto M., Saar E., Tucker D. L., Müller V., Heinämäki P., Allam S. S., 2003, A&A, 405, 425
Eke V. R., Baugh C. M., Cole S., Frenk C. S., Norberg P., Peacock J. A., Baldry I. K., Bland-Hawthorn, J. et al., 2004a, MNRAS, 348, 866
Eke V. R., Frenk C. S., Baugh C. M., Cole S., Norberg P., Peacock J. A., Baldry I. K., Bland-Hawthorn, J. et al., 2004b, MNRAS, 355, 769
Erdoğdu P., Lahav O., Zaroubi S., Efstathiou G., Moody S., Peacock J. A., Colless M., Baldry, I. K. et al., 2004, MNRAS, 352, 939
Forero-Romero J. E., Hoffman Y., Gottlöber S., Klypin A., Yepes G., 2009, MNRAS, 396, 1815
Gaztañaga E., Norberg P., Baugh C. M., Croton D. J., 2005, MNRAS, 364, 620
Geller M. J., Huchra J. P., 1989, Science, 246, 897
Gilbank D. G., Balogh M. L., 2008, MNRAS, 385, L116
Gott J. R. I., Jurić M., Schlegel D., Hoyle F., Vogeley M., Tegmark M., Bahcall N., Brinkmann J., 2005, ApJ, 624, 463
Hahn O., Porciani C., Carollo C. M., Dekel A., 2007, MNRAS, 375, 489
Jackson J. C., 1972, MNRAS, 156, 1P
Kaiser N., 1987, MNRAS, 227, 1
Kim H., Baugh C. M., Cole S., Frenk C. S., Benson A. J., 2009, MNRAS, 400, 1527
Nichol R. C., Sheth R. K., Suto Y., Gray A. J., Kayo I., Wechsler R. H., Marin F., Kulkarni, G. et al., 2006, MNRAS, 368, 1507
Norberg P., Cole S., Baugh C. M., Frenk C. S., Baldry I., Bland-Hawthorn J., Bridges T., Cannon, R. et al., 2002, MNRAS, 336, 907
Novikov D., Colombi S., Doré O., 2006, MNRAS, 366, 1201
Park C. et al., 2005, ApJ, 633, 11
Pimbblet K. A., Drinkwater M. J., 2004, MNRAS, 347, 137
Pimbblet K. A., Drinkwater M. J., Hawkrigg M. C., 2004, MNRAS, 354, L61
Praton E. A., Melott A. L., McKee M. Q., 1997, ApJ, 479, L15+
Schaap W. E., van de Weygaert R., 2000, A&A, 363, L29
Sousbie T., Colombi S., Pichon C., 2009, MNRAS, 393, 457
Sousbie T., Pichon C., Colombi S., Novikov D., Pogosyan D., 2008a, MNRAS, 383, 1655
Sousbie T., Pichon C., Courtois H., Colombi S., Novikov D., 2008b, ApJ, 672, L1
Stoica R. S., Martínez V. J., Mateu J., Saar E., 2005, A&A, 434, 423
Stoica R. S., Martínez V. J., Saar E., 2010, A&A, 510, A38+
Yaryura Y. C., Baugh C. M., Angulo R. E., 2010, arXiv:astro-ph/0007281
York D. G., Adelman J., Anderson, Jr. J. E., Anderson S. F., Annis J., Bahcall N. A., Bakken J. A., Barkhouser, R. et al., 2000, AJ, 120, 1579
Zehavi I., Blanton M. R., Frieman J. A., Weinberg D. H., Mo H. J., Strauss M. A., Anderson S. F., Annis, J. et al., 2002, ApJ, 571, 172
Zhang Y., Yang X., Faltenbacher A., Springel V., Lin W., Wang H., 2009, ApJ, 706, 747